PROGRAM BOOK

# 제5회 국제인권심포지엄

# The 5th International Human Rights Symposium

## 혐오표현의 확산, 어떻게 대응할 것인가? – 해외 사례로 보는 제도적 대안

## How to Respond to the Spread of Hate Speech?
## - Institutional Countermeasures Learning from International Cases

2019년 8월 28일 (수) 10:00 ~ 16:30 | 서울 서대문구 연세대학교 동문회관 그랜드볼룸

10am - 4:30pm, Wednesday, August 28, 2019
Grand Ballroom, Yonsei University Alumni Association Bldg., Seodaemun-gu, Seoul

Ministry of Justice

# 제5회 국제인권심포지엄

The 5th International Human Rights Symposium

혐오표현의 확산, 어떻게 대응할 것인가? – 해외 사례로 보는 제도적 대안

How to Respond to the Spread of Hate Speech?
- Institutional Countermeasures Learning from International Cases

## Table of Contents


| | | |
|---|---|---|
| 프로그램 | Program | 04 |
| 세션1 | Session 1 | 07 |
| 세션2 | Session 2 | 115 |

## 프로그램

| 시간 | 내용 | |
|---|---|---|
| 09:30 - 10:00 | 등록 및 안내 | |
| 10:00 - 10:10 | 개회사 | **박상기** 법무부 장관 |
| | 축사 | James Lynch 유엔 난민기구 한국대표부 대표 |
| 세션 1. 혐오표현의 국내외 현황과 관련제도 | | |
| 10:10 - 12:00 | 사회 | **홍성수** 숙명여자대학교 교수 |
| | 발제1 | 혐오표현과 EU의 정책들<br>**Joëlle Hivonnet** 주한 EU대표부 차석대사 |
| | 발제2 | 국내 혐오표현의 유형과 대응방안<br>**류민희** 공익인권변호사모임 희망법 변호사 |
| | 토론1 | **이수연** 한국여성정책연구원 선임연구원 |
| | 토론2 | **Kimberley Barker** University of Stirling 교수 |
| 12:00 - 13:30 | 오찬 | |
| 세션 2. 혐오표현과 차별, 그 제도적 대안은? | | |
| 13:30 - 15:20 | 사회 | **김민정** 한국외국어대학교 교수 |
| | 발제1 | 사례발표 1 – 관련 법제 도입과정의 장애 및 극복 방안<br>**Evan Balgord** Canadian Anti-Hate Network 이사 |
| | 발제2 | 사례발표 2 – 관련 법제 도입과정의 장애 및 극복 방안<br>**김창호** 일본 변호사 |
| | 발제3 | 혐오표현 관련 해외 규제 제도의 시사점과 국내 도입 검토<br>**이주영** 서울대학교 인권센터 전문위원 |
| | 토론1 | **김지혜** 강릉원주대학교 교수 |
| | 토론2 | **김정학** 국가인권위원회 혐오차별대응기획단 팀장 |
| 15:20 - 15:30 | 휴식 | |
| 종합토론 | | |
| 15:30 - 16:30 | 사회 | **홍성수** 숙명여자대학교 교수 |

## Program

| Time | Content | |
|---|---|---|
| 09:30 - 10:00 | Registration and Introduction | |
| 10:00 - 10:10 | Welcome Remarks | **Sang-ki Park** Minister of Justice, Republic of Korea |
| | Congratulatory Speech | **James Lynch** Representative, UNHCR Korea |
| Session 1. Current Contours of Hate Speech and Regulations | | |
| 10:10 - 12:00 | Moderator | **Sung-soo Hong** Prof. of Sookmyung Women's University |
| | Presentation 1 | EU policies on hate speech<br>**Joëlle Hivonnet** Chargé d'Affaires of the EU Delegation to the Republic of Korea |
| | Presentation 2 | Types and countermeasures of hate speech in South Korea<br>**Min-hee Ryu** Lawyer of Korean Lawyers for Public Interest and Human Rights |
| | Discussion 1 | **Soo-yeon Lee** Senior Research Fellow, Korean Women's Development Institute |
| | Discussion 2 | **Kimberley Barker** Lecturer of University of Stirling |
| 12:00 - 13:30 | Luncheon | |
| Session 2. Hate Speech and Discrimination: What are the Policy Alternatives? | | |
| 13:30 - 15:20 | Moderator | **Min-jung Kim** Prof. Of Hankuk University of Foreign Studie |
| | Presentation 1 | International Case 1<br>**Evan Balgord** Executive Director of Canadian Anti-Hate Network |
| | Presentation 2 | International Case 2<br>**Chang-ho Kim** Lawyer in Japan |
| | Presentation 3 | Implications of the international regulations and institutions against hate speech and ways to introduce them in Korea<br>**Dr. Joo-Young Lee** SNU Human Rights Center |
| | Discussion 1 | **Ji-hye Kim** Prof. of Gangneung-Wonju National University |
| | Discussion 2 | **Cheong-Hak Kim** Hate Speech Project Team Manager of National Human Rights Commission of Korea |
| 15:20 - 15:30 | Break | |
| General Discussion | | |
| 15:30 - 16:30 | Moderator | **Sung-soo Hong** Prof. of Sookmyung Women's University |

제5회 국제인권심포지엄

The 5th International Human Rights Symposium

혐오표현의 확산, 어떻게 대응할 것인가? – 해외 사례로 보는 제도적 대안

How to Respond to the Spread of Hate Speech?
- Institutional Countermeasures Learning from International Cases

# Session 1

## 혐오표현의 국내외 현황과 관련 제도
## Current Contours of Hate Speech and Regulations

### 사회자 | Moderator

**홍성수 Sung-soo Hong**
숙명여자대학교 교수
Prof. of Sookmyung Women's University

### 발제자 | Speakers

**Joëlle Hivonnet**
주한 EU대표부 차석대사
Chargé d'Affaires of the EU Delegation to the Republic of Korea

**Katharine Gelber**
University of Queensland 교수
Prof. of University of Queensland

**류민희 Min-hee Ryu**
공익인권변호사모임 희망법 변호사
Lawyer of Korean Lawyers for Public Interest and Human Rights

### 토론자 | Discussants

**이수연 Soo-yeon Lee**
한국여성정책연구원 선임연구원
Senior Research Fellow, Korean Women's Development Institute

**Kimberley Barker**
University of Stirling교수
Lecturer of University of Stirling

제5회 국제인권심포지엄

The 5th International Human Rights Symposium

혐오표현의 확산, 어떻게 대응할 것인가? – 해외 사례로 보는 제도적 대안

How to Respond to the Spread of Hate Speech?
- Institutional Countermeasures Learning from International Cases

# Session 1

## 혐오표현과 EU의 정책들
## EU Policies on hate speech


발제자 | Speaker

**Joëlle Hivonnet**

주한 EU대표부 차석대사

Chargé d'Affaires of the EU Delegation to the Republic of Korea

오늘 이 자리에 초대해주셔서 감사합니다. 유럽연합(EU) 내에서도 혐오표현을 중요한 주제로 다루고 있으며 동 사안을 해결하기 위하여 다각적으로 노력하고 있습니다.

배경에 대하여 간략하게 설명한 후 혐오범죄의 기록, EU 행동강령, 주요 IT 기업의 사례를 들어 불법적인 온라인상의 혐오표현에 어떻게 대응하고 있는지 말씀 드리겠습니다.

**배경**

인권보호 및 증진은 유럽연합 창설의 핵심 가치이자 외교정책의 중심입니다. 유럽연합은 내부 법 제정을 통해 성, 인종, 민족, 종교, 신념, 장애, 나이, 성 정체성에 기반한 차별을 척결하고 있습니다. 성평등을 증진하기 위한 법도 시행하고 있습니다. 외부적으로는 시민단체를 지원하여 인권 및 시민자유 증진을 도모하는 한편 불관용과 맞서 싸우고 있습니다.

혐오표현과 혐오범죄를 해결하기 위해서는 차별에 대한 개념을 정립하는 것이 중요합니다. 혐오표현과 혐오범죄는 동일한 기폭제로 인해 발생하며, 일부 사람들을 2 급시민으로 취급하고 그들의 존엄성과 권리를 존중하지 않습니다. 유럽연합은 2000 년부터 시행한 자체 차별방지법을 통해 위 문제들을 해결하기 위해 노력해왔습니다.

혐오표현과 혐오범죄는 구체적으로 나타나는 현상으로 특정한 행동을 동반하며, 차별과도 연관되어 있습니다. 이러한 이유로 유럽연합과 회원국은 혐오표현과 혐오범죄에 대응하기 위한 법을 제정하고 피해자에 대한 지원서비스를 도입해왔습니다. 추후에 확인하실 수 있겠지만, 혐오표현과 범죄에 대한 명확한 정의가 필요하고 이를 범죄의 한 분류로 인식해야 하기 때문에, 위의 피해자 지원 서비스는 혐오범죄를 신고하고 기록하는 데에 중요한 역할을 합니다.

용어 정의에 대해 말씀 드리겠습니다. 혐오표현은 특정 그룹의 사람들에 대한 혐오, 차별 또는 적대감을 퍼뜨리고, 유발하며, 조장하거나 정당화하는 공개적인 표현을 뜻합니다. 혐오표현은 불관용적인 분위기를 조성하여 대상 그룹에 대한 공격을 더욱 용이하게 만듭니다. 혐오표현의 보편적인 목적은 특정 사회 그룹의 인간 존엄성과 가치를 훼손하는 것입니다. 일부 사례에서는 혐오표현이 범죄로 간주될 수 있습니다.

혐오범죄는 특정 그룹의 실제 또는 사회적으로 인식되어 있는 특성을 대상으로 하는 모든 형태의 범죄를 말합니다. 혐오범죄는 신체·정신적 위해, 협박, 손괴, 위협 및 폭력, 강간, 살인 등 다양한 형태로 나타날 수 있습니다. 일부 사례에서 상황을 악화시키는 것도 혐오범죄에 해당할 수 있습니다.



유럽연합에서 정의하는 혐오표현과 혐오범죄의 피해자는 여성, 유대교인, 이슬람교인, 집시, 성소수자(LGBT)입니다.

위의 정의가 빛을 발하려면 피해자가 혐오로 인한 희롱이나 폭력을 경찰에 신고하고 경찰관이 위의 사고를 혐오범죄로 기록해야 합니다.

**유럽연합의 혐오범죄 기록과 자료수집 관행**

작년에 발간된 유럽연합기본권청(FRA) 보고서는 이 현상을 정확하게 이해하고 관계당국이 혐오표현과 범죄에 효과적으로 대응하기 위해서는 데이터 수집이 중요하며, 그 이전에 혐오범죄의 신고와 기록이 선행되어야 한다고 강조했습니다. 이러한 조치가 취해지면 혐오범죄의 피해자들의 사법 접근성 역시 개선될 것입니다.

FRA 보고서 주요 내용은 다음과 같습니다.

- 혐오범죄는 대부분 경찰에 신고되지 않습니다. 이러한 상황에서 FRA 는 혐오범죄를 특정하는 질문을 포함한 설문조사를 시행하여 혐오범죄를 제대로 조명하고 신고되지 않은 범죄를 수면 위로 끌어올릴 필요가 있다고 말합니다.
- 위의 이유로 데이터는 산재되어 있습니다. 범죄 동기 및 유형 등 혐오범죄에 대한 상세하고 세분화된 데이터를 수집하여 경찰이 효과적으로 대응하고 있는지 모니터링 할 필요가 있습니다.
- 유럽연합 군사참모부 19 개국이 공개한 혐오범죄 데이터 중, 15 개국만 동 데이터를 편향 동기로 세분화하였습니다.
- 시민사회단체와 교육, 피해자 지원, 혐오 범죄 기록 분야에서의 협력을 통해 기록방법의 비교성 및 호환성을 개선하는 것이 중요합니다.
- 조직문화는 피해자와 공동체에 대한 법 집행 대응에 매우 큰 영향을 끼칩니다. 법 집행 최고지도부가 혐오범죄에 대한 대응 의지를 다지고 그 필요성을 인정해야 하며, 혐오범죄 기록에 대한 중요성을 인식해야 합니다. 모든 조직원들과 이러한 의지를 공유하고 실천해야 합니다.
- 협력, 투명성, 책임성에 기반한 치안문화를 형성하여 경찰에 대한 대중의 신뢰를 제고하고 피해자들의 범죄 신고를 독려해야 합니다.

사견을 덧붙이자면, 저는 벨기에나 스페인 등 일부 유럽연합 군사참모부 국가에서 성별 별로 데이터를 세분화한 반면, FRA 의 데이터가 그렇지 않다는 점에 놀랐습니다.

다음으로 온라인상의 혐오표현에 대하여 말씀 드리겠습니다.

**온라인상의 혐오표현에 대한 유럽연합 행동강령**

2016 년 5 월 유럽집행위원회와 4 개 주요 IT 업체는(페이스북, 마이크로소프트, 트위터, 유튜브→ 이후 구글, 인스타그램, 스냅챗, 데일리모션, 투데이, 웨비디아 합류) 온라인상의 혐오표현에 대응하기 위한 행동강령을 발표했습니다. 동 행동강령 이행 현황 검토에 따르면 다양한 분야에서 성과가 있었습니다.

동 행동강령에 서명한 모든 IT 업체는 현재 서비스 약관, 규정 또는 커뮤니티 이용 기준을 마련하여 사용자가 개인 그룹에 대한 폭력이나 혐오를 조장하는 콘텐츠를 게재하는 것을 금지하고 있습니다. 해당 콘텐츠를 감시하고 검토하는 직원 수를 확충하였으며, 페이스북의 경우 전 세계 1 만 5 천명의 직원을 동원하여 모든 유형의 콘텐츠를 검토하고 있다고 밝혔습니다.

IT 업체들은 평균적으로 위험 콘텐츠의 89%를 24 시간 내에 감지하고 있습니다. 정기적인 교육을 실시하는 한편, 콘텐츠 검토 담당 팀에 교육과 지원을 제공합니다.

IT 업체들은 이른바 '신뢰할 수 있는 기수(trusted flaggers; 위험 콘텐츠 모니터링 인력)' 네트워크를 대폭 확대하였으며, 정기적인 소통을 통해 혐오표현의 국가별 특성에 대한 이해를 제고하고 있다고 보고했습니다. 행동강령('17)을 체결한 이후 첫 해에 페이스북은 66 개 유럽연합 비정부기구를 '신뢰할 수 있는 기수'로 동원하였다고 밝혔으며, 트위터 역시 21 개 유럽연합 회원국 내 40 개 비정부기구와 협업했다고 밝혔습니다.

동 행동강령에 동의한 모든 IT 기업은 국가적 차원의 연락소를 설치하여 유관 당국과 국가 차원에서의 긴밀한 소통을 이어갔다. 이를 통해 당국이 온·오프라인상의 불법 혐오표현 작성자를 기소해야 하는 법 제도를 보완할 수 있습니다.

IT 업체들은 이전에 비해 어떻게 온라인상의 혐오표현에 대한 신고를 대처해왔는지에 대한 정보를 보다 올바른 방식으로 공유하고 있으며, 그 예로는 페이스북, 트위터, 유튜브가 발표한 투명성보고서를 들 수 있습니다.

동 행동강령 출범 이전에는 사용자들이 혐오표현 콘텐츠 신고 시 IT 기업으로부터 관련 답변을 받는 경우는 거의 없었습니다. 이에 더해 신고 또는 플래깅(flagging) 기능은 때로 사용자 친화적이지 않았습니다. 위 기업들의 성과와 실적에는 차이가 있지만, 평균적으로 2/3 의 혐오표현 신고 건에 대하여 처리 결과와 실행 조치 등을 상세히 설명하는 체계적인 답변을 제공하였습니다.

2018 년부터 인스타그램, 구글플러스, 스냅챗, 데일리모션, 제욱스비디오닷컴 (jeuxvideo.com) 5 개 기업이 동 행동강령에 동참하고 있습니다. 이로써 총 9 개 기업이 참여하고 있으며, 혐오 콘텐츠에 영향을 받을 수 있는 유럽연합 온라인 플랫폼 시장의



96%의 점유율을 보유하게 되었습니다. 행동강령은 산업표준이 되었으며 다양한 플랫폼, 대기업 및 중소기업의 참여를 독려하고 있습니다. 동 행동 강령은 혐오표현이라는 난제를 해결하는 데에 있어 모든 이해관계자들을 한데 모으는 구심점 역할을 합니다. 하지만 트래픽 용량 등 해결하기 쉽지 않은 과제들이 남아있습니다.

**결론**

마지막으로 인공지능을 활용하여 온라인상의 혐오표현을 탐지하는 방법이 있다는 것을 말씀 드리고 싶습니다. 금년 브뤼셀에서 유럽규제센터 주관으로 개최된 세미나에서 동 주제를 다룬 바 있습니다. 많은 사람들이 인적자원 부족으로 데이터 관리가 어렵고 콘텐츠가 혐오표현에 해당하는지 결정 시 개별 사용자 세분화가 힘들다는 문제를 인공지능을 통해 해결할 수 있을지 주목합니다. 실제로 인공지능을 개발하여 혐오표현을 탐지하고 삭제할 수 있는 가능성은 분명히 존재합니다만, 현재까지 진행된 연구에 따르면 학습하는 기계가 노출 또는 성활동을 나타내는 이미지 패턴을 파악할 수는 있지만 혐오표현의 문맥과 변화하는 형태를 이해하는 시스템을 개발하기는 훨씬 더 어렵다고 합니다. 정확도는 개선할 수 있는 부분이지만, 오늘날 이러한 기술의 활용을 통해 문맥을 파악할 때 포용성을 제대로 나타내지 못하는 위험이 있을 수 있습니다.

좋은 소식은 혐오표현에 대한 인식이 제고되고 있으며 정계에서도 아직 승리하지 못한 혐오표현과 범죄와의 전쟁에서 승리하겠다는 강한 의지를 드러내고 있다는 점입니다.

경청해주셔서 감사합니다.

Thank you for inviting me today because hate speech is a topic of growing importance in the EU and one that we are trying to address in a comprehensive fashion.

First of all, I would like to give a little bit of background and then I will focus on the challenge of hate crime recording and then on the Code of conduct that was elaborated by the EU and some major IT companies to counter illegal hate speech online.

***Background***

The protection and promotion of Human Rights is a core value that the European Union is founded upon and key to its foreign policy. Internally, there is EU legislation in place to combat discrimination based on sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation. There is also legislation in place to promote gender equality. In its external action, the EU also seeks to support civil society organisations that promote human rights and civil liberties and combat intolerance.

I want to stress how important it is to have an operational concept of discrimination before you can address the issue of hate speech and crime because they proceed from the same triggers, which lead to treating some people as second class citizens with little respect for their dignity and their rights. In the EU we have had anti-discrimination legislation in place since 2000, addressing all the grounds I have mentioned previously.

Hate speech and hate crime are specific phenomena that are clearly linked to discrimination and, as such, they require specific action. This is why the EU and its Member States have introduced laws against hate speech, hate crime and support services for the victims. As you will see later, the role of such services is crucial for the reporting and recording of hate crimes because hate speech and hate crime need to be recognised as a specific category of offense and crime, with a definition.

In terms of definition, hate speech refers to public expressions that spread, incite, promote or justify hatred, discrimination or hostility towards a specific group of persons. They contribute to a general climate of intolerance, which, in turn, makes attacks more probable against those given groups. In general, hate speech aims to undermine the dignity and value of a human being belonging to a particular social group. In some cases, hate speech can be deemed to be a crime.

Hate crime is any form of crime targeting people because of their actual or perceived belonging to a particular group. The crimes can manifest in a variety of forms: physical and psychological intimidation, blackmail, property damage, aggression and violence, rape, and murder. In some cases, they constitute aggravating circumstances.

In the EU, the victims of hate speech and hate crime are women, Jewish, Muslim, Roma or from a sexual minority (LGBT).

Such definitions are only of any use if the victims report hate-motivated harassment and violence to the police, and if police officers record such incidents as hate crimes.

***Hate crime recording and data collection practice across the EU***

In a report published last year, the European Fundamental Rights Agency (FRA) underlined the importance of data collection and, before that, the importance of the reporting and recording of hate crime in order to have an accurate picture of the phenomenon, which will, in turn, allow the authorities to address hate speech and hate crime effectively. This will also help ensuring a better access to justice for the victims of hate crime.

The main findings of the FRA report are the following.

- The majority of hate crimes are not reported to the police. In such a situation, FRA insists on the need to conduct surveys that include hate crime-specific questions, in order to shed light on the dark figure of crime *i.e.* unreported hate crime.
- As a result, data is patchy. The collection of detailed and disaggregated data on hate crime - at minimum, by bias motivation and by type of crime – is necessary to monitor the effectiveness of the police response. Of the 19 EUMS that publish data on recorded hate crime, only 15 disaggregate these data by different bias motivations.
- Effective engagement with civil society organisations in the areas of training, victim support and hate crime recording is essential. This would also help improve the comparability and compatibility of recording methodologies.
- The prevailing organisational culture heavily influences law enforcement responses to victims and communities. The highest levels of the law enforcement hierarchy must embrace and acknowledge their commitment to countering hate crime and the importance of properly recording such crime. Then they need to communicate this commitment to the rank and file and implement it.
- Creating a culture of policing based on cooperation, transparency and accountability could improve public confidence in the police and encourage victims to report crime.

I will also venture to make a personal comment. I was shocked to find out that sex does not feature in the categories identified by the FRA although it is clearly identified as a category by some EUMS, for example Belgium or Spain.

Now I would like to turn to the specific challenge of online hate speech.

***EU Code of Conduct on countering hate speech online***

In May 2016, the European Commission and four major IT companies (Facebook, Microsoft, Twitter and YouTube – (joined later by Google, Instagram, Snapchat, Dailymotion, Today and Webedia) presented a Code of Conduct on countering hate speech online.

The latest review of implementation of the Code of Conduct indicated good results in many areas.

All IT companies that signed the Code now have terms of service, rules or community standards prohibiting users from posting content inciting violence or hatred against groups of individuals. They have also significantly increased the number of employees monitoring and reviewing the content. Facebook reports having a global network of about 15,000 people working on all types of content review.

On average, IT companies are now assessing 89% of flagged content within 24 hours.

IT companies reportedly hold regular training, and provide coaching and support for their team of content reviewers

IT companies reported a considerable extension of their network of so-called "trusted flaggers" in Europe and are engaging on a regular basis with them to improve the understanding of national specificities of hate speech. In the first year after the signature of the Code of conduct (2017), Facebook reported it had taken 66 EU NGOs on board as "trusted flaggers" and Twitter 40 NGOs in 21 EU countries.

All IT companies that subscribed to the Code of Conduct have established national points of contact to facilitate contact with the relevant competent authorities at national level. This complements the legal framework that requires authorities to prosecute the authors of illegal hate speech offences - whether online or offline.

IT companies are now much better in presenting information about how they have followed up reports of hate speech online, for example in transparency reports published by Facebook, Twitter and YouTube.

Before the Code of Conduct was launched, users rarely received a response by IT companies when they notified hate speech content. In addition, the reporting or flagging function was often not very user-friendly. While there is scope for progress and the performance of the companies differs in this respect, now on average around two thirds of the notifications receive a systematic response detailing the outcomes and measures taken.



Since 2018, five new companies joined the Code of Conduct: Instagram, Google+, Snapchat, Dailymotion and jeuxvideo.com. This brings the total number of companies that are part of the Code of Conduct to nine, with a coverage of 96% of the EU market share of online platforms which may be affected by hateful content. The Code has become an industry standard, attracting a diverse range of platforms and both big and smaller businesses.

The Code of Conduct is at the centre of a strategy that seeks to bring together all stakeholders in addressing the challenge of hate speech but is facing severe challenges due, in particular, to the volume of traffic.

**Conclusion**

Therefore, by way of conclusion, I would like to point to the potential use of artificial intelligence to identify hate speech online, following a seminar held in Brussels earlier this year and organised by the Centre on Regulation in Europe.

Increasingly, interest centres on whether AI could provide a solution to the problem of not having enough human moderators to deal with increasing amounts of data and the issue of individual human discretion in deciding whether content amounts to hate speech.

Indeed there are clearly possibilities in developing AI to automatically detect and delete hate speech. However, research so far suggests that while learning machines can identify patterns in images of nudity or sexual activity, it is more difficult to develop systems that can understand the context and changing form of hate speech. Until accuracy can be improved, the use of such technology at this time would present dangers of over or under inclusiveness in identifying context.

What is important is that there is increasing awareness of the phenomenon and a clear political will to address it but the battle against hate speech and hate crime has yet to be won.

I thank you for your attention!

제5회 국제인권심포지엄

The 5th International Human Rights Symposium

혐오표현의 확산, 어떻게 대응할 것인가? – 해외 사례로 보는 제도적 대안

How to Respond to the Spread of Hate Speech?
- Institutional Countermeasures Learning from International Cases

# Session 1

## 혐오표현 대응에 관한 국제적 기준과 현황
## International and Australian approaches to hate speech laws


발제자 | Speaker

**Katharine Gelber**

University of Queensland 교수

Prof. of University of Queensland

# 혐오표현 규제법에 대한 국제적 접근
# : 호주의 사례에서 얻은 교훈

**서론**

오늘 호주의 혐오표현 규제법 시행 방식에 대해 중점적으로 알아보고자 한다. 호주는 형법이 아닌 민법이 더 강조되는 국가이기 때문이다. 따라서 본 방식은 문제에 대한 잠재적인 정책방안을 검토하는 유용한 수단이 될 것이며, 이는 국제적으로 드문 사례이기도 하다.

전 세계적으로 다양한 혐오표현 규제법이 존재하고 있는데 (eg Brown 2015: 19-48) 여기에는 집단명예훼손, 부정적 고정관념화, 증오 조장, 공공질서 위반, 집단학살 부정, 불법행위, 혐오범죄 등에 대한 것들이 포함되어 있다. 시간상 관계로 여기 나열한 모든 법에 대해 살펴볼 수는 없겠지만 한 가지 언급하고 싶은 점은 많은 국가들이 혐오표현을 형법으로 규제하고 있다는 사실이다. 즉 법을 어겼을 시 기소나 구속이 가능하다는 이야기다. 즉, 법 규정의 엄격성으로 인해 아주 극단적인 경우가 아니라면 혐오표현 관련 범죄는 법망을 빠져나가거나 이에 대한 법적대응이 어려워지는 경향이 있다.

**다음은 수정 후 발췌 부분, 출처:** Gelber, K & McNamara, L 2015 'The effects of civil hate speech laws: lessons from Australia', Law and Society Review 49(3): 631-664. (DOI: 10.1111/lasr.12152).

호주는 혐오표현 규제를 위한 다양한 규제 방안을 고안하였는데, 이는 관련 분야에 새로운 정책을 도입하고자 하는 사법관할에 큰 도움을 줄 수 있을 것이다. 이에 오늘 호주의 연방 정부, 주정부, 특별자치구 차원에서 시행되고 있는 관련 법 규정의 핵심 내용들에 대해 설명하고자 한다. 고소가 이루어지는 방식, 가능한 법적 배상 유형, 혐오표현 규제법의 효과에 대해 집중적으로 다룰 계획이다. 또한 혐오표현의 정의, 고소의 권리가 인정되는 자, 법이 보호하지 않는 영역 등 주요 특징들에 관해서도 설명하는 시간을 가지고자 한다.

1

**호주의 법**

호주는 6 개의 주와 2 개의 준주 (準州) 로 이루어진 연방국가이다. 북쪽 준주를 제외한 모든 관할권이 혐오표현 규제법을 제정하였다. 호주는 이와 관련하여 인종주의적 혐오표현에 대한 형사적 및 민사적 규정을 제정하였으며 많은 관할권에서는 이 외에도 성 정체성, 종교, 트랜스젠더 지위, 장애, HIV/AIDS 감염 여부 등에 관한 규제를 하고 있다. (표 1 참조)

*(표 1. 혐오표현에 대한 민사적 규제 도입 연대표)*

| 관할권 | 도입 년도 | 추가 규제 내용 |
|---|---|---|
| 뉴사우스웨일즈주 | 1989 | 인종 – 1989<br>동성애 – 1993<br>HIV/AIDS – 1994<br>트랜스젠더 – 1996 |
| 호주 수도 특별구 | 1991 | 인종 – 1991<br>HIV/AIDS – 2004<br>트랜스섹슈얼 – 2004<br>섹슈얼리티- 2004 |
| 커먼웰스주 | 1995 | 인종 - 1995 |
| 사우스오스트레일리아주 | 1996 | 인종 - 1996 |
| 태즈매니아주 | 1998 | 인종 – 1998<br>종교 – 1998<br>섹슈얼리티– 1998<br>장애 - 1998 |
| 퀸즈랜드주 | 2001 | 인종 – 2001<br>종교 – 2001<br>섹슈얼리티– 2002<br>성 정체성- 2002 |
| 빅토리아주 | 2001 | 인종 – 2001<br>종교 - 2001 |

2

1980 년대 후반에서 1990 년대 초반 최초로 제정된 혐오표현 규제법의 등장 배경으로는 두 가지 주요 원인을 꼽을 수 있다. 첫째, 뉴사우스웨일즈주의 '내셔널 액션'이나 웨스턴오스트레일리아주의 '호주 민족주의 운동'과 같은 우파 조직들에 의한 악성 혐오표현의 난무와 이에 대한 우려를 들 수 있다. (McNamara 2002: 121, 222–25) 둘째, 인권및기회평등위원회가 발표한 인종차별적 폭력행위 실태조사에 따르면 소수인종과 토착민 공동체를 향한 학대, 위협, 공포심조장, 차별, 폭력 등의 인종차별이 충격적인 수준인 것으로 나타났음을 알 수 있다. (AHREOC 1991).

혐오표현 규제법의 도입이 표현의 자유에 영향을 줄 것이라는 우려가 있음에도 불구하고 현존하는 법이 공공연한 인종차별적 표현행위를 제재하거나 막는데 부적절하다고 판단되었을 뿐 아니라 호주의 다문화주의와 평등과 비차별원칙을 관철시키고자 하는 의지를 실현하고자, 국회는 위와 같은 법의 도입을 적극 추진하였다.

1990 년대와 2000 년대에 걸쳐 혐오표현의 규제대상이 확대되었는데 그 이유로는 두 가지를 들 수 있다: 1) 관할권 내 주요 인종차별금지 규정의 광범위한 변화로 인해 (예: 태즈메이니아주), 혹은 2) 만연하는 혐오표현에 대한 지역 사회의 우려를 잠식시키기 위해 위와 같은 확대가 있었다. 예를 들어, 뉴사우스웨일즈주는 동성애자들에 대한 폭력 확산에 대응의 일환으로 1990 년대 초, 동성애 혐오표현을 규제 범위 안에 포함시켰다.

또한, 호주의 다자간 인권 협약 비준으로 국제적인 인권 보호 정서가 확산되었으며 이는 국회의 입법을 촉진시키는 역할을 하였다. (McNamara 2002: 36).

혐오 표현 규제법 도입의 제안 표현의 자유에 대한 우려의 목소리도 있었다.
혐오표현 규제법 도입이 제안되었을 때 표현의 자유에 대한 우려의 목소리도 있었다. 몇몇 반대론자들은 비방금지법을 두고 '위험한 법' 혹은 '사상의 자유를 억압하려는 시도'라고 이야기하였다.[1] 그렇지만 대부분의 사람들은 법안의 초안이 잘 만들어지기만 한다면 한 쪽엔 표현의 자유, 다른 한쪽엔 인간 존엄성 보장의 권리와 차별 받지 않을 권리를 두고 양 쪽의 균형을 잘 맞출 수 있을 것이라 생각한다. 입법자들은 예외조항을 둠으로써 균형을 맞출 수 있었다. 곧이어 이에 대한 상세한 설명을 할 예정이다.[2]

비록 호주 수도 특별구를 비롯한 국가의 모든 주에서 형법이 시행되고 있긴 하지만 그 적용례는 거의 찾아볼 수 없다. 거의 30년 동안 오직 6건의 성공적인 기소가 이루어졌을

[1] McNamara, *Regulating Racism*, 36.
[2] For example *Anti-Discrimination Act 1977* (NSW), s 20C; *Racial Discrimination Act 1975* (Cth), s 18D; McNamara, *Regulating Racism*, 127-130.





뿐이다[3]. 뉴사우스웨일즈주[4], 퀸즈랜드주[5], 사우스오스트레일리아주[6]의 형법은 다음과 같은 행위를 금지한다.

- 혐오 선동, 심한 모욕, 극심한 조롱
- 그리고 이와 동시에 신체적 위해 혹은 신체적 위해의 협박, 개인이나 집단 혹은 그들의 재산에 대해 물리적 손해를 가하도록 타인을 선동하는 행위

빅토리아주[7]에서는 형사법상 다음과 같은 행위를 금지한다

- 혐오선동, 협박 혹은 협박을 가하도록 타인을 선동, 개인이나 그들의 재산에 대한 물리적 손해를 가하는 행위

호주의 수도 특별구[8]에서는 고의적인 혹은 주의의무 위반에 의한 혐오 선동, 심한 모욕, 극심한 조롱이 수반되는 협박 행위를 형사법상 금지하고 있다.

비교해 볼만한 흥미로운 점은, 웨스턴오스트레일리아주는 형사상으로만 혐오표현을 규제하고 있다는 것인데, 다음과 같은 행위로서 이중 범죄를 야기한다. (행위 자체 혹은 의도하지 않은 행위에 기반)

* 인종간 적개심 선동 혹은 인종 괴롭힘을 퍼뜨리기 위한 자료의 보유
* 인종적으로 괴롭히는 행위
* 인종적 괴롭힘을 야기하는 전시행위를 위한 자료의 보유

해당 주에서는 3 건의 성공적인 기소가 있었다. 여기에는 2005 년의 인종차별주의적 자료의 보유로 인한 기소(Gelber 2007:8), 2006 년의 "인종적 괴롭힘의 가능성이 있는 행위"에 대한 유죄판결 (ODPP WA 2011), 그리고 2009 년의 "인종간 적개심 혹은 인종

[3] There have been 3 prosecutions in Western Australia (Katharine Gelber and Luke McNamara, "The Effects of Civil Hate Speech Laws: Lessons from Australia," *Law and Society Review* 49, no. 3 (2015): 635); 2 prosecutions in Queensland: one in 2015 involving racial abuse to which the accused pleaded guilty and received a suspended sentence (Anti-Discrimination Commission Queensland, *Submission to Religious Freedom Review* (Feb 2018), https://www.pmc.gov.au/sites/default/files/religious-freedom-submissions/14531.docx; Australian Associated Press, "Teen's Racist Brisbane Train Rant Disgusting: Magistrate," *Brisbane Times* (14 Sep 2015), http://www.brisbanetimes.com.au/queensland/teens-racist-brisbane-train-rant-disgusting-magistrate-20150914-gjm2e7.html), and a second in which a transgender female was verbally threatened physical harm (private correspondence to author from the Anti-Discrimination Commission, Queensland, 19 April 2018); and 1 in Victoria (James Oaten, "Far-right Nationalists Found Guilty of Inciting Serious Contempt for Muslims after Mock Beheading Video," *ABC News* (5 Sep 2017), http://www.abc.net.au/news/2017-09-05/three-men-found-guilty-of-inciting-serious-contempt-for-muslims/8874804.

[4] *Crimes Act 1900*, s93Z
[5] *Anti-Discrimination Act* 1991 (Qld), s 131A.
[6] *Racial Vilification Act* 1996 (SA), s 4.

[7] *Racial and Religious Tolerance Act* 2001 (Vic), ss 24, 25.
[8] *Discrimination Act* 1991 (ACT), s 67.

괴롭힘을 선동할 의도가 있는 행위"와 "인종적 괴롭힘의 가능성이 있는 행위"에 대한 기소로서 2012 년 항소에 성공하지 못한 케이스가 있다.

호주에서는 규제에 있어 민법이 실질적인 역할을 한다. 대부분의 소가 민사분쟁으로 이루어진다. (Gelber and McNamara 2014a).

호주의 혐오표현 규제 관련 국내법[9]은 다음과 같이 기술되어 있다.

1. 사적인 행위 이외의 것으로서 아래와 같은 행위를 하는 것은 불법이다.
a. 상황에 상관 없이, 타인이나 집단에게 불쾌감, 모욕감, 굴욕감을 주거나 위협을 가할 가능성이 상당한 행위
b. 타인 혹은 집단의 일부 혹은 전부에 행해지는 행위의 기준이 그들의 인종, 국적, 종족 출신지 등에 의한 경우

비록 피해 판단의 기준이 상대적으로 낮아 보일 수 있겠으나, 법원은 충족되어야 할 기준점을 "단순 경미한 것이 아닌 극심하고 심각한 영향을 주는 행위"[10]로 정립하였다. 표현의 자유 보호를 위해 신의성실의 원칙에 따른 예술적, 학술적, 과학적, 그리고 언론과 관련된 행위는 예외 규정으로 본다.[11]

뉴사우스웨일즈주는 1989 년에 최초로 혐오표현 규제법을 제정하였다.[12] 그 내용은 다음과 같다.

S20C (1) 개인이 공공행위로서, 인종을 이유로 타인 혹은 집단을 특정 지어 행하는 혐오 선동 행위, 심한 모욕, 극심한 조롱은 불법이다.

뉴사우스웨일즈주는 다음과 같은 예외 규정을 두고 있다.

a. 세부항목 (1)에 언급된 공공 행위에 대한 공정 보도, 혹은
b. 명예훼손에 관한 소송절차 진행 중 절대적 면책사유의 대상이 될 수 있는 의사소통

[9] *Racial Discrimination Act 1975* (Cth), s18C.
[10] *Eatock v. Bolt* (2011) 283 ALR 505, 561.
[11] *Racial Discrimination Act* 1975 (Cth) s 18D
[12] *Anti-Discrimination Act* 1977 (NSW), ss 20B-D, 38R-38T, 49ZS-49ZT, 49ZXA- 49ZXC

5

c. 교육, 예술, 과학 혹은 연구 목적이나 토론, 논쟁 등 공익을 위한 기타 목적이 있는 경우, 다른 행위나 문제에 대한 설명을 하는 경우, 합리적이고 신의성실에 맞게 행해진 공공 행위

민법은 혐오표현 사건 발생시 인권 당국에 서면으로 소장을 제출하도록 요구하고 있다. (뉴사우스웨일즈주의 반차별위원회나 호주 연방정부의 인권위원회에 제소 가능) 해당 기관은 기소 내용을 조사하여 실제로 비방이 이루어졌는지를 검토하고 원고와 피고 사이의 기밀 조정을 실시한다. 제시될 수 있는 배상에는 행위의 즉시 중지, 사과, 철회 의사 게재 (신문 등의 매체에) 혹은 직장 내에서의 교육을 들 수 있다.

기소자는 기소 행위를 종결한 후 각 재판소 (주 또는 특별자치구 소속) 혹은 연방법 하의 연방법원에서 민사 절차 진행에 착수할 수 있다. 혐오표현으로 기소된 사건 중 정식 판결까지 가는 비율은 2 퍼센트 미만이며 그 중 절반만이 유죄판결을 받게 된다. (Gelber and McNamara 2014a: 314) 재판소 혹은 법원에 의해 제소된 행위가 불법적인 혐오표현임이 판단될 시 위 기관은 피고에게 사과, 중지 명령, 손해배상금 지급[13], 시정 통지 발행 등을 명령할 수 있다. 제소가 유효하게 이루어지기 위해서는 혐오표현 행위가 반드시 공연성이 인정되어야 하며, 이 말인 즉 공공이 혐오발언을 들을 수 있다는 가능성이 상당 부분 예측 되어야한다. (Chapman and Kelly 2005: 207–8, 210–13).

호주 정부가 시행하는 혐오발언의 민사적 규제의 중심에는 피해자들이 있다. 혐오표현의 피해자인 개인 혹은 단체의 대표자들만이 본 법의 주체가 된다. 이러한 사실은 증오범죄를 조사하고 기소해야 할 국가의 책임과 극명히 모순되는 것으로, (Jenness and Grattet 2001) 호주의 혐오발언 규제법에 의하면 어떤 국가 기관도 기소나 소송을 진행시킬 권한을 가지고 있지 않다. (Gelber and McNamara 2014a: 307).

**혐오표현 금지법은 얼마나 잘 작동할까?**

*피해자들을 위한 적절한 구제가 이루어졌는가?*

호주의 혐오표현 규제법에 있어 우선적으로 생각해보아야 할 사항은 배상문제이다. 혐오표현의 피해자들은 성공적인 소 제기를 통해 문제의 개선과 만족스러운 결과를 얻을 수 있었는가?

첫째, 어느 해를 보더라도 소 제기 건수는 비교적 적은 편임을 알 수 있다. 2010년을

[13] The legislative limit for damages in a vilification case is $100,000, however, orders are typically $10,000 or less (Gelber and McNamara 2014a: 314).

6

기준으로 지난 10년간 전국적으로 제기된 기소 건수는 매년 최대 342건에서 최소 165건 사이였다. 이는 호주의 인구가 약 2천만명인 점, 대부분의 관할권이 비방금지법을 채택하고 있으며 법이 규제하는 대상도 다양하다는 점 등을 고려해 볼 때 상대적으로 적은 수치라는 것을 알 수 있다.

둘째, 검증을 위한 새로운 법안이 도입된 직후, 다른 때보다 더 많은 수의 기소가 이루어졌음을 알 수 있다. 초반의 이러한 추세는 법 도입 후 어느 정도의 시간이 지났을 때와 다른 양상이다. 예를 들어, 2004년-2005년에는 기소 건수가 전년도에 비해 크게 증가했음을 알 수 있다. 이 중 거의 절반 정도가 한 관할권, 즉 태즈매니아주에서 이루어졌으며 이는 모두 국가의 비방금지법 도입 직후에 일어났다.

법의 도입 후 처음 몇 년 동안 법을 이용하고자 하는 사람이 많아지는 이유는 아마 새롭게 제정된 법에 대한 인식이 높아진 상태에서 그 효용성과 적용 가능성을 확인해 보고자 하는 욕구가 더해졌기 때문일 것이다. 호주유대인평의회 (ECAJ)의 선출직 임원 제레미 존스는 인종관련 혐오표현 규제법을 도입하는 데 중요한 역할을 하였으며, 이를 통해 반유대주의 문제를 다루고자 하였다. 처음 법이 도입되었을 때 존스는 인터뷰를 통해 그들의 조직이 '사람들이 개인의 힘으로 문제를 해결함에 있어 가장 무력함을 느낄 때는 언제 인지, 어느 시점에서 이들을 위한 행위를 취해야 하는지'에 대해 살펴보았다고 언급하였다. (Jones 2013) 이러한 사건들은 계속해서 다뤄졌고, 이에 따라 법의 적용을 위해 필요한 주요 요소들을 명확하게 함으로써 사건이 행동으로 조치될 수 있는 최소한의 임계점을 포함하여, 인터넷 상 자료가 법조항에 포섭된다는 점, 홀로코스트 부정 역시 금지된다는 점을 판단할 수 있었다. 그 후 지역 사회는 기타 다른 사건들을 다룰 때 있어 이와 동일한 판단 기준을 사용할 수 있게 된 것이다.
이러한 판단 기준은 사람들의 비방 행위를 막고자 그들을 설득하는 도구로서 쓰였다. 비록 재판소 법원의 공식적인 재판으로 해결되는 경우는 2퍼센트 미만이고, 이에 따라 공개적으로 발표되는 판단 역시 적더라도 말이다. 만약 사건이 기밀 조정으로 해결된다면 이를 교육 자료로 사용할 기회는 매우 제한적이 된다. 인권 당국은 일부 익명 사례들을 연례 보고서로 제출하지만 자세한 내용은 공개하지 않는다.

전반적으로, 혐오표현의 민사적 규제는 두 가지 차원에서 해결책을 제공하고 있다. 첫번째는 소가 제기될 수 있다는 점, 그리고 경우에 따라 유리한 결과를 얻어낼 수 있다는 점이다. 정부의 권한을 이용할 수 있다는 사실은 혐오표현이 법에 어긋난다는 것을 증명해주며 이는 그 자체로 중요한 의미를 갖는다. 이를 통해 지역 사회는 차별로부터 법적 보호를 받기 위해 필요한 지식을 쌓을 수 있기 때문이다. 두번째는 법의 교육적 역할에 관한 것이다. 이러한 역할에는 혐오표현을 사용하고자 하는 사람을 기존 선례를 통해 설득하는 직접적 방법도 포함된다. 형사적 방법과 달리 혐오 표현의



민사적 방법은 더 다양한 종류의 표현 행위를 다룰 수 있기 때문에 이는 특히나 더 중요한 부분이라고 볼 수 있다. 그러나 이러한 해결책은 지역사회에는 상당한 수준의 혐오표현이 지속적으로 존재한다는 점, 기소를 통해 문제를 끝까지 해결하는 과정에서 오는 부담감이 클 수 있다는 점, 혜택이 지역 사회 전반에 고르게 돌아가지 않는 다는 점에 그 한계가 있다.

*표현의 수정?*

우리는 각 관할의 신문에 개제된, 편집자 앞으로 온 6,612 통의 편지에 대한 질적 연구를 시행하였다. (Gelber and McNamara 2014b) 혐오표현 규제법으로 인해 일반적인 표현의 사용에 있어 질적인 변화가 보이는지 판단하기 위함이었다.

연구 결과 첫째, 작성자들은 혐오표현 규제법에 대한 지식이 있으며 이러한 지식이 그들이 사용하는 표현에 직접적으로 영향을 준다는 사실을 알게 되었다. 이는 규제와 관련된 용어들이 사용되었다는 점, 그리고 이러한 용어의 등장 시기가 법의 도입과 시행 시기와 일치한다는 점으로부터 알 수 있었다.

둘째, 우리는 시간이 지남에 따라 성 정체성 기반의 편견을 표현하는 언어가 지속적으로 변화한다는 사실을 발견하였다. 1994 년에는 성 정체성 기반 편견을 표현함에 있어 "동성애자의 구역질나는 행동"과 같은 표현이 사용되었다. 이와는 대조적으로 2004 년의 편지에는 "생활 방식의 선택 (성적 파트너의 선택)" 과 같은 표현들이 사용되었다.

호주 토착민에 대한 편견을 표현하는데 있어 쓰이는 언어 역시 지속적이진 않지만 확실한 변화를 보였다. 1990 년대 초반에는 "미개한" 등의 표현이 쓰였지만 1990 년대 중반에 들어서는 편견을 나타냄에 있어 주로 "호주원주민들을 위한 특별법"이나 도난당한 세대 "신화" 등의 표현이 쓰였다.

최근의 이민자들에 대한 편견을 표현하는 언어에 있어서는 일관적인 변화를 찾지 못하였으며, "그들이 왔던 곳으로 이민자들을 다시 돌려보내라," "유색인종범죄," "시끄러운 소수자들," 등의 표현이 쓰였다. 망명 신청자들을 이야기함에 있어서는 "불법 이민자," "테러리스트" "초대받지 못한 불법 침입자" "새치기하는 사람" 등의 표현이 쓰였다.

세번째로 우리는 편지 전수조사를 통해 시간이 갈수록 편견을 나타내는 표현의 작지만 유의미한 감소를 발견할 수 있었다. 기간에 따라 편지를 세 기간으로 나눈 경우, 1992-1997 년 사이에 쓰여진 "편견적인" 편지의 비율은 33.86 퍼센트, 1998-2003 년 사이에는

29.08 퍼센트, 2004-2009 년 사이에는 28.54 퍼센트의 수치를 보였다. 이러한 편견적 표현의 감소는 유익한 결과라고 보여 진다. 물론 예를 들어, 동성결혼의 권리에 반대하는 사람들도 있겠지만 혐오표현 규제법의 목표 중 하나는 현재 논란이 있는 공공 정책 사안에 대한 논쟁 자체를 차단하는 것이 아닌 비방이 없는 건전한 토론의 장을 도모하고자 함에 있다. 우리의 연구는 (예를 들어) 동성 결혼에 대한 반대의 입장 혹은 찬성의 입장을 보는 것이 아니라, 그 입장을 표명하는데 있어 혐오표현을 사용 하였는가의 여부를 판단함에 그 목적이 있다.

*교육적이고 상징적인 효과?*

우리는 법이 교육적인 역할을 수행하는데 있어 두 가지 방법을 이용한다는 결론을 낼 수 있었다. 1) 계속되는 비방을 방지하기 위해 가해자에게 법원이 내린 불법행위 판결에 대해 이야기함으로써 지역사회를 위한 직접적이고 의식적인 효과를 볼 수 있다. 그리고 2) 덜 직접적이고 수량화하기도 어렵긴 하지만, 신문에 게재된 편지 중 편견적인 내용이 점차 줄어드는 것에서부터 알 수 있었듯이 법에는 교육적인 효과가 있음이 증명되었다.

우리는 또한 해당 지역 사회 구성원을 인터뷰하였다. 혐오표현 규제법의 중요성에 대한 의견을 묻는 질문을 받았을 때, 구성원들은 이에 대한 엄청난 지지를 보여주었다. 그들은 해당 법이 그 자체로 취약한 지역사회를 지지하는 역할을 한다고 믿었다. 인터뷰에 응했던 사람들은 단순히 "법이 그곳에 존재하고 있음을 아는 것" 이 중요하다고 설명했으며 법을 통해 무엇이 "용인되지 않는 행위"인지에 대한 기준을 세울 수 있다고 이야기했다.

토착민 인터뷰 대상자들은 특히 호주의 혐오표현 규제법이 모든 사람에게 공통으로 적용되는 기준을 세우는데 유용한 역할을 할 것이라고 보았다.

*냉각효과*

혐오표현 규제법에 대해 지속적으로 제기되는 주장이 바로 냉각효과이다. 즉, 규제로 인해 공개 토론이 사라질 것이라는 것이다. 그렇지만 혐오표현을 규제하고 있는 나라 어디에서도 이에 대한 증거는 찾아볼 수 없다.

신문에 게재된 편지 조사 건에서도 밝혀진 바이지만 지난 25 년동안 공개 토론이 줄어들었다는 증거는 나오지 않았다. 호주 토착민들의 토지에 대한 권리, 동성 결혼, 망명 신청자의 처우 등 광범위한 쟁점들을 주제로 한 다양한 토론이 활발히 진행되고 있다. 연구 결과를 보면 시간이 지남에 따라 편견적 표현이 꾸준히 존재해 왔다. 또한,



분석 결과 무절제한 언어 사용 방식으로부터의 변화가 포착되었다는 사실만으로 냉각효과의 주장에 힘이 실리는 것도 아니다. 이러한 주장의 핵심은 관점과 견해에 대한 침묵에서 오는 우려이기 때문이다.

따라서 우리는 바람직한 효과와 바람직하지 못한 효과를 구분 짓는다. 혐오표현 규제법은 개개인이 자신의 관점과 견해를 표현하는데 사용하는 언어에 영향을 주고자 고안되었다. (바람직한 효과) 그렇지만 특정 주제에 관하여는 "언급 금지" 할 수밖에 없도록 만들어지기도 했다. (바람직하지 못한 효과) 우리의 연구 결과 냉각 효과의 위험성은 아직 그 영향이 입증되지 않았다고 보인다. 호주인들은 광범위한 정책 사안들에 대한 그들의 관점과 견해를 활발히 이야기할 준비가 언제든 되어있다.

*순교자의 양산?*

혐오표현 규제법이 순교자를 양산한다는 주장은 무엇일까? 호주의 언론인 앤드류 볼트가 연방정부의 인종차별금지법에 저촉되는 행위를 한 사건이 있었는데 이 때문에 혐오표현 금지법이 순교자를 양산할 수 있다는 주장에 힘이 실리게 되었다. 2011년 볼트의 인종차별금지법의 위반이 인정된 후 사건이 조직적으로 재구성되었고 이는 언론의 담론을 지배했다. 언론은 볼트를 피해자로 만들었고 혐오표현 규제법을 두고 표현의 자유를 침해하는 것처럼 묘사했다.

하지만 여기에는 양 자간에 균형이 필요하다. 지난 20년 이상의 시간동안 순교효과의 계기가 될 만한 민사소송 사례는 없었다. 예를 들어 극악무도한 홀로코스트 부정론자 프레드릭 토빅은 볼트 사건 몇 년 전, 마찬가지로 인종차별금지법을 위반한 적이 있었는데 당시 토빅은 순교자 지위를 취하려고 했었다. 연방 법원은 토빅에게 그의 웹사트에서 홀로코스트를 부정하는 자료를 삭제하라고 명령했지만 그는 이를 거절했다. 그 결과 24건의 법정모독으로 기소되었고 결국 죄가 인정되어 3개월의 징역형을 선고받았다. (Akerman 2009) 공공 담론에서의 시도는 그의 악명높은 지위를 더욱 공고히 해줬고, 동시에 이는 애초에 왜 혐오표현 규제법이 존재해야 하는지에 대한 이유를 정확하게 보여주는 대표적 사례가 되었다. 호주의 혐오표현 규제법에서 주목해야 할 두 가지 특징이 있다. 첫째, 대부분의 법 위반은 기밀 조정으로 다루어지며 2퍼센트 미만의 사건만이 법원이나 재판소의 결정을 받게 된다는 점에서 공공의 영역으로 이어지는 사건의 수는 많지 않다. 따라서 재판의 결과로 순교의 길을 걷게 되는 경우는 드물다. 둘째, 본 법은 민사상 해결에 의지하는 비율이 압도적으로 높기 때문에 순교효과의 근거라고 할 수 있는 형사 제재를 받게 될 확률은 미미하다.

**호주의 경험에서 얻은 교훈**

1. 우리가 사용할 수 있는 정의는 다양하게 존재한다.
2. 민법을 강조하면 혐오표현 규제법에 대한 비판을 줄일 수 있다.
3. 법의 예외규정을 잘 만든다면 공공의 논쟁을 줄이고 법에 의한 표현의 자유 침해를 최소화할 수 있다.
4. 혐오표현 규제법은 사람들이 대중 토론에 참여하는 방식을 바꾸도록 고안된 것이며 특정 토론 주제를 제한시키고자 하는 것이 아니다.
5. 혐오표현 규제법으로 얻는 이득은 상당하며, 특히 해당 지역 개개인은 물론 사회 공동체 전반에 긍정적 영향을 가져다준다.
6. 혐오발언법으로 인한 부정적 영향은 미미한 수준이다.



# International approaches to hate speech laws : lessons from Australia

## Introduction

I will focus today on the way Australian has implemented hate speech laws, because there is an emphasis in Australia on the civil law (and not the criminal law). This is a very useful way to examining potential policy approaches to the problem, and it is unusual internationally.

There is a wide variety of hate speech laws in existence internationally (eg Brown 2015: 19-48), which includes laws that target group defamation, negative stereotyping, incitement to hatred, public order offences, denial of genocide, tort actions, and hate crimes. I do not have time to provide an overview of all these laws here, but I want to note that many countries around the world prioritise a criminal law response to hate speech. This means that a violation of the law can result in prosecution and imprisonment. Because of the seriousness of these provisions, they tend to be used in the most extreme cases and therefore much 'hate speech' is not captured by, or responded to with, these kinds of laws.

**The following is excerpted, with amendments, from:**

**Gelber, K & McNamara, L 2015 'The effects of civil hate speech laws: lessons from Australia', *Law and Society Review* 49(3): 631-664. (DOI: 10.1111/lasr.12152).**

Australia has developed a variety of approaches to regulating hate speech, which can be very instructive for those jurisdictions considering wanting to develop new policy in this area. Today I want to outline some of the core features of the provisions operating in Australia at federal and state/territory level. I will focus on the mechanisms by which complaints can be made, the types of remedies that are available, and the effects of the hate speech laws. I hope to highlight key features including definitions of hate speech, who has grounds for complaint, and what hate speech laws do *not* do.

## Australia's laws

Australia is a federation with six states and two self-governing Territories. All jurisdictions except the Northern Territory have enacted hate speech laws. The Australian approach to hate speech regulation has involved the enactment of both criminal and civil provisions against racist hate speech, with many jurisdictions including other grounds such as sexuality, religion, transgender status, disability, and HIV/AIDS status (see Table 1).

*Table 1. Chronology of the Introduction of Civil Hate Speech Laws in Australia*

| Jurisdiction | Civil law first enacted | Ground/s added |
|---|---|---|
| New South Wales | 1989 | Race – 1989<br>Homosexuality – 1993<br>HIV/AIDS – 1994<br>Transgender – 1996 |
| Australian Capital Territory | 1991 | Race – 1991<br>HIV/AIDS – 2004<br>Transsexuality – 2004<br>Sexuality - 2004 |
| Commonwealth | 1995 | Race - 1995 |
| South Australia | 1996 | Race - 1996 |
| Tasmania | 1998 | Race – 1998<br>Religion – 1998<br>Sexuality – 1998<br>Disability - 1998 |
| Queensland | 2001 | Race – 2001<br>Religion – 2001<br>Sexuality – 2002<br>Gender identify - 2002 |
| Victoria | 2001 | Race – 2001<br>Religion - 2001 |

There were two major drivers for the enactment of the first hate speech laws in the late 1980s and early 1990s. First, there were concerns about the virulent hate speech being circulated by right wing organizations (such as National Action in New South Wales and the Australian National Movement in Western Australia) (McNamara 2002: 121, 222–25). Second, in 1991, the National Inquiry into Racist Violence conducted by the Human Rights and Equal Opportunity Commission documented disturbing levels of racism directed at ethnic minority and Indigenous communities, which manifested in harassment, intimidation, fear, discrimination, and violence (AHREOC 1991).

Although the introduction of hate speech laws raised concerns about their implications for freedom of expression, legislatures were motivated to act because existing laws were seen to be inadequate to sanction and deter public expressions of racism, and out of a determination to mark Australia's commitment to multiculturalism and principles of equality and non-discrimination (ALRC 1992; McNamara 2002: 18–20).

The extension of protection to other grounds through the 1990s and 2000s was for two reasons: 1) either part of a wider updating of the jurisdiction's primary antidiscrimination statute (e.g., in Tasmania), or 2) a response to local concerns about the prevalence of hate speech. For example, New South Wales extended hate speech laws to cover hate speech on the ground of homosexuality in the early 1990s in response to a reported increase in the prevalence of violence against homosexuals.



The international human rights sentiments generated by Australia's ratification of relevant multilateral human rights treaties also played a role in prompting legislatures to act (McNamara 2002: 36).

There were free speech concerns aired in response to proposals to introduce hate speech laws. Some opponents described anti-vilification laws as "dangerous," an "attempt to stifle freedom of thought."[1] But the majority view was that such laws, when appropriately drafted, achieved the right balance between the right to freedom of speech on the one hand, and the right to a life of dignity and free from discriminatory harms on the other. Parliamentarians achieved this balance by including exemptions in the laws, which I will discuss in more detail in a moment.[2]

Although criminal laws have been implemented in all the states and the Australian Capital Territory, they are rarely used. There have only been six successful prosecutions in nearly 30 years of their operation.[3] The criminal laws in New South Wales,[4] Queensland,[5] and South Australia,[6] prohibit conduct that:

- incites hatred, serious contempt or severe ridicule, and
- simultaneously involves physical harm or the threat of harm, or inciting others to threaten physical harm toward a person, a group of persons, or their property.

Victoria[7] criminally prohibits conduct that:

- incites hatred and threatens, or incites others to threaten, physical harm toward a person or their property.

In the Australian Capital Territory,[8] the criminal law prohibits threatening conduct that intentionally and recklessly involves the incitement of hatred, serious contempt or severe ridicule.

As an interesting counterpoint, Western Australia possesses only criminal hate speech laws, which create two-tiered offences (based on the existence or not of intent) of conduct that:

- incites racial animosity or racial harassment,
- possession of material for dissemination that incites racial animosity or racial

[1] McNamara, *Regulating Racism*, 36.
[2] For example *Anti-Discrimination Act 1977* (NSW), s 20C; *Racial Discrimination Act 1975* (Cth), s 18D; McNamara, *Regulating Racism*, 127-130.
[3] There have been 3 prosecutions in Western Australia (Katharine Gelber and Luke McNamara, "The Effects of Civil Hate Speech Laws: Lessons from Australia," *Law and Society Review* 49, no. 3 (2015): 635); 2 prosecutions in Queensland: one in 2015 involving racial abuse to which the accused pleaded guilty and received a suspended sentence (Anti-Discrimination Commission Queensland, *Submission to Religious Freedom Review* (Feb 2018), https://www.pmc.gov.au/sites/default/files/religious-freedom-submissions/14531.docx; Australian Associated Press, "Teen's Racist Brisbane Train Rant Disgusting: Magistrate," *Brisbane Times* (14 Sep 2015), http://www.brisbanetimes.com.au/queensland/teens-racist-brisbane-train-rant-disgusting-magistrate-20150914-gjm2e7.html), and a second in which a transgender female was verbally threatened physical harm (private correspondence to author from the Anti-Discrimination Commission, Queensland, 19 April 2018); and 1 in Victoria (James Oaten, "Far-right Nationalists Found Guilty of Inciting Serious Contempt for Muslims after Mock Beheading Video," *ABC News* (5 Sep 2017), http://www.abc.net.au/news/2017-09-05/three-men-found-guilty-of-inciting-serious-contempt-for-muslims/8874804.
[4] *Crimes Act 1900*, s93Z
[5] *Anti-Discrimination Act* 1991 (Qld), s 131A.
[6] *Racial Vilification Act* 1996 (SA), s 4.
[7] *Racial and Religious Tolerance Act* 2001 (Vic), ss 24, 25.
[8] *Discrimination Act* 1991 (ACT), s 67.

harassment,
- conduct that racially harasses, and
- possession of material for display that racially harasses.

There have been three successful prosecutions in that state: one for possession of racist material in 2005 (Gelber 2007: 8); a guilty plea to "conduct likely to racially harass" in 2006 (ODPP WA 2011); and a prosecution for "conduct intended to incite racial animosity or racist harassment" and for "conduct likely to racially harass" in 2009, which was unsuccessfully appealed in 2012.

The civil laws carry the practical regulatory burden in Australia – they are used in the vast majority of complaints (Gelber and McNamara 2014a).

Australia's national hate speech law[9] relevantly states:

1. It is unlawful for a person to do an act, otherwise than in private, if:

    a. the act is reasonably likely, in all the circumstances, to offend, insult, humiliate or intimidate another person or a group of people; and

    b. the act is done because of the race, color or national or ethnic origin of the other person or of some or all of the people in the group.

While the harm threshold may look as though it is relatively low, the courts have established that the standard to be met is conduct that has "profound and serious effects, not to be likened to mere slights."[10] To protect free speech, exemptions apply to artistic, academic, scientific, and journalistic conduct, done in good faith.[11]

New South Wales was the first jurisdiction to enact a hate speech law in 1989:[12] Its law states:

> s20C (1) It is unlawful for a person, by a public act, to incite hatred toward, serious contempt for, or severe ridicule of, a person or group of persons on the ground of the race of the person or members of the group.

In NSW, there is an exemption for:

a. a fair report of a public act referred to in subsection (1), or

b. a communication ... that would be subject to a defence of absolute privilege ... in proceedings for defamation, or

c. a public act, done reasonably and in good faith, for academic, artistic, scientific or research purposes or for other purposes in the public interest, including discussion or debate about and expositions of any act or matter.

[9] *Racial Discrimination Act 1975* (Cth), s18C.
[10] *Eatock v. Bolt* (2011) 283 ALR 505, 561.
[11] *Racial Discrimination Act* 1975 (Cth) s 18D.
[12] *Anti-Discrimination Act* 1977 (NSW), ss 20B-D, 38R-38T, 49ZS-49ZT, 49ZXA- 49ZXC



The civil laws require a person who believes an incident of hate speech has occurred to lodge a complaint in writing with a human rights authority (e.g., the Anti-Discrimination Board in New South Wales, or the Australian Human Rights Commission under federal law). The authority investigates the complaint to ascertain whether vilification has occurred, and seeks to conciliate a confidential settlement between the complainant and respondent. The kinds of remedies that can be provided include an agreement to stop, an apology, or an agreement to publish a retraction (eg in a newspaper), or to conduct an educational campaign in a workplace.

A complainant may terminate a complaint and commence civil proceedings in a tribunal (in a state or Territory) or the Federal Court (under federal law). Less than 2 percent of hate speech complaints are formally adjudicated and half of those produce findings that the conduct complained of was unlawful (Gelber and McNamara 2014a: 314). If the tribunal/court determines that the conduct in question is unlawful hate speech, it can order an apology, an order to stop, the payment of damages,[13] or the publication of a corrective notice. For a complaint to be valid the conduct must have occurred in public, which means that it needed to be reasonably foreseeable that a member of the public could have heard the speech (Chapman and Kelly 2005: 207–8, 210–13).

The locus of enforcement under Australia's regime of civil hate speech laws rests with the victims. Legislation can only be invoked by an individual or representative organization from the group that has been subjected to hate speech. In sharp contradistinction to the responsibility of the state to investigate and prosecute alleged hate crimes (Jenness and Grattet 2001), under Australia's civil hate speech laws no state agency has the authority to initiate a complaint or to pursue litigation (Gelber and McNamara 2014a: 307).

**How well do the laws work?**

*A remedy for targeted groups?*

The first question to consider is whether Australia's hate speech laws provide a remedy. Are targets are able successfully to lodge complaints for incidents of hate speech and achieve an outcome that ameliorates its effects?

First, the number of complaints in any given year is relatively modest. In the decade up to 2010 the total number of complaints nationally fluctuated from a high of 342 to a low of 165 per year. These are relatively modest numbers of complaints, given the size of the Australian population at approximately 20 million, and the extent of anti-vilification laws that cover most jurisdictions and a variety of grounds.

Second, there appears to be a trend shortly after new legislation is introduced to test it out, as evidenced by relatively higher numbers of complaints compared with later periods. For example, the year 2004–2005 shows a significant increase in the numbers of complaints compared with the previous year. Nearly half of these complaints were in one jurisdiction—Tasmania—and occurred shortly after the introduction of that state's anti- vilification laws.

[13] The legislative limit for damages in a vilification case is $100,000, however, orders are typically $10,000 or less (Gelber and McNamara 2014a: 314).

It may be that the higher use of the law in the first few years after it is introduced is due to a heightened awareness of the newly enacted legislation, combined with a desire to test its utility and application. Jeremy Jones, an elected official of the Executive Council of Australian Jewry (ECAJ), has been instrumental in invoking racist hate speech laws to address antisemitism. Jones told us in interview that when the laws were first introduced, their organization looked at, "where do people feel most unable to respond as individuals, and where are we getting people saying we have to do something?" (Jones 2013) These cases were pursued and clarification of key aspects of the law's operation obtained, including the threshold required for an incident to be actionable, that material on the internet was covered by the provisions, and that Holocaust denial was prohibited. Subsequently, the community was able to use those judgments in combatting other incidents:

The judgments were used as a tool of advocacy to convince people not to engage in vilification. This was the case even though less than 2 percent of matters are resolved by formal adjudication in a tribunal or court, and, therefore, produce judgments that are released publicly. Where matters are resolved by confidential conciliation, there is very limited opportunity to use these outcomes for educational purposes. Human rights authorities report on some anonymised case studies in their annual reports, but do not release details.

Overall, civil hate speech laws are providing a remedy, in two senses. The first is that complaints can be lodged and in some cases a favorable outcome obtained. The ability to have a governmental authority validate the message that hate speech breaches the law is important in and of itself, because it provides targeted communities with the knowledge that the law can assist in protecting them from discrimination. The second is in terms of the laws' educative role. This educative role includes directly using precedents to dissuade hate speakers. Given the ability of the civil hate speech model to target a wider range of expressive conduct than a purely criminal model would permit, this is particularly important. The remedies are, however, limited in the senses that there are persistent, significant levels of hate speech in the community, the burden on complainants in seeing a complaint through can be high, and there is an uneven distribution of the benefits among target communities.

*A modification of speech?*

We conducted a qualitative document analysis of 6,612 letters to the editor published in newspapers in each jurisdiction (Gelber and McNamara 2014b) to determine whether the quality of moderated speech had changed as a result of hate speech laws.

Our analysis showed, first, that writers of letters to the editor demonstrate knowledge of hate speech laws, and a connection between those laws and the expression they are using. This is indicated both by the presence of terms related to the laws, and the timing of the emergence of those terms to coincide with the introduction, or expansion, of hate speech laws.

Second, we found a sustained shift over time in the language used to express sexuality-based prejudice. In 1994, letters expressing sexuality-based prejudice used



terms like "sick act of homosexuality." In contrast, in 2004 letters expressing sexuality-based prejudice used terms such as "lifestyle choice."

We found a discernible, but less sustained, shift in language used to express prejudice toward Indigenous Australians. In the early 1990s terms such as "uncivilised" were in prejudicial letters. By the mid 1990s prejudice was primarily conveyed by referring to "special laws for Aborigines," and the stolen generations "myth."

We found no consistent shift in language used to express prejudice toward recent migrants, with expressions including "send migrants back where they came from," "ethnic crimes," "noisy minorities," and descriptions of asylum seekers as "illegal immigrants," "terrorists," "uninvited intruders," and "queue-jumpers."

A third finding is that, in the total population of letters analysed, there was a modest but significant reduction in the expression of prejudice over time. When the letters are divided into three equal time periods, the proportion of "prejudicial" letters published in 1992–1997 was 33.86 percent, in 1998–2003 the figure was 29.08 percent and in 2004–2009 the figure was 28.54 percent. This reduction in the expression of prejudice is a beneficial outcome. While some might still oppose the right, for example, of same sex couples to marry, one of the aims of hate speech laws is not to shut down debate on controversial issues of public policy, but to assist in generating a debate that does not vilify. What was captured by our analysis is not the expression of views opposing or supporting (for example) same sex marriage, but whether in expressing their views, the writer engaged in hate speech.

*An educative and symbolic effect?*

We have already concluded that there have been two ways in which the laws play an educative role: 1) the direct and conscious use of prior judgments in community advocacy to curb ongoing vilification by telling the perpetrators that the court has stated that what they are doing is unlawful, and 2) a less direct and harder to quantify, educative effect evidenced by the reduction in the proportion of prejudicial letters published in newspapers.

We also interviewed target community members. When they were asked if they thought hate speech laws were important, they expressed overwhelming support for them. They believed that the laws were useful as a statement in support of vulnerable communities. Interviewees described it as important simply to "know they're there" and that they set a standard for what's "not acceptable." Indigenous interviewees particularly saw Australia's hate speech laws as useful in setting a standard against which all people should be held to account.

So the legal form of hate speech laws may be less important than the fact of their existence.

*A 'chilling effect'*

A regular claim against hate speech laws is that they will have a chilling effect, which means they will shut down public debate. But there is no evidence that this is the case in any country that has hate speech laws.

Our analysis of letters to the editor revealed no evidence that public discourse has been diminished over the past 25 years. Robust debates are being held on a broad range of issues including the land rights of Indigenous Australians, same-sex marriage, and the treatment of asylum-seekers. Our analysis revealed the continued expression of prejudice over time. The fact that we detected a shift away from more intemperate styles of language cannot be said to support the chilling effect claim because at the heart of this claim is a concern about the *silencing* of views and opinion.

Therefore, we distinguish between desirable and undesirable effects. Hate speech laws are designed to influence the terms in which individuals express their views in public (desirable effect), however, they are not designed to make certain topics "off limits" (undesirable effect). Our research suggests that the risk of a chilling effect has not been substantiated. Australians are willing to express robust views on a broad range of policy issues.

*The creation of martyrs?*

What of the claim that hate speech laws produce martyrs? In Australia there is a story of journalist Andrew Bolt's encounter with federal racial hatred laws which does lend some support to the claim that hate speech laws can produce martyrs. In 2011 after Bolt was found to have breached federal racial hatred law, an orchestrated reconstruction of the decision dominated media discourse. That coverage positioned Bolt as a victim, and portrayed hate speech laws as incursions into free speech.

Yet a sense of proportion is required here. No other case in over two decades of civil litigation has triggered a comparable martyr effect. For example, recalcitrant Holocaust denier Frederick Toben attempted to adopt a martyr position when he was found to have breached the same federal racial hatred law years earlier.[39] His refusal to abide by orders of the Federal Court to remove Holocaust denial material from his website resulted in 24 contempt of court findings and, ultimately, a 3 month jail term for contempt of court (Akerman 2009). However, in public discourse this attempt served to consolidate his infamy and status as a powerful illustration of precisely why hate speech laws should be enacted in the first place (Aston 2014; Richardson 2014). Two distinctive features of Australia's hate speech laws are noteworthy. First, given, that most transgressions of the law are addressed in confidential conciliation, with less than 2 percent resulting in court or tribunal decisions that enter the public domain, opportunities for martyrdom are rare. Second, because the laws rely overwhelmingly on *civil* remedies, they tend not to produce the criminal sanctions on which the claimed martyr effect is based.

**Lessons from the Australian experience**

1. There is a variety of definitions available.
2. An emphasis on civil law reduces many of the criticisms of hate speech law.
3. Well drafted exemptions in the law preserve public debate and ensure the law does not infringe too far on freedom of speech.
4. Hate speech laws are designed to change how people engage in public debate, not to make some topics off limits for public debate.
5. The benefits of hate speech laws are considerable, especially to target communities, and through that to the community as a whole.
6. There are negligible negative effects from having hate speech laws.

THE UNIVERSITY OF QUEENSLAND AUSTRALIA | CREATE CHANGE

# 혐오표현 규제법에 대한 국제적 접근: 호주의 사례에서 얻은 교훈

Professor Katharine Gelber
Head of School
School of Political Science and International Studies
University of Queensland
k.gelber@uq.edu.au

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

## *혐오표현에 대한 민사적 규제 도입 연대표*

| 관할권 | 도입 년도 | 추가 규제 내용 |
|---|---|---|
| 뉴사우스웨일즈주 | 1989 | 인종 – 1989<br>동성애 – 1993<br>HIV/AIDS – 1994<br>트랜스젠더 – 1996 |
| 호주 수도 특별구 | 1991 | 인종 – 1991<br>HIV/AIDS – 2004<br>트랜스섹슈얼– 2004<br>섹슈얼리티 – 2004 |
| 커먼웰스주 | 1995 | 인종 – 1995 |
| 사우스오스트레일리아주 | 1996 | 인종 – 1996 |
| 태즈매니아주 | 1998 | 인종 – 1998<br>종교 – 1998<br>섹슈얼리티 – 1998<br>장애 – 1998 |
| 퀸즈랜드주 | 2001 | 인종 – 2001<br>종교 – 2001<br>섹슈얼리티 – 2002<br>성 정체성– 2002 |
| 빅토리아주 | 2001 | 인종 – 2001<br>종교 – 2001 |

CRICOS code 00025B 2

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

## 혐오표현 규제법 도입 이유

1. 현존하는 법이 공공연한 인종차별적 표현행위를 제재하거나 막는데 부적절하다고 판단
2. 호주의 다문화주의 실현에 대한 의지
3. 평등과 비차별에 대한 의지
4. 국제 인권법상의 의무
5. 표현의 자유와 차별적 행위의 위해로부터 자유로울 권리의 균형을 맞추기 위해

➢ 예외규정:
- ✓ 공정보도
- ✓ 예술, 교육, 과학 연구 목적, 혹은 '선의의' 공개토론회
- ✓ 실제 사건에서 민법에 비중을 둔 경우

CRICOS code 00025B 3

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

## 형사상 혐오표현 규제법

➢ 호주의 모든 관할권에 존재하는 것이 아님

➢ 거의 사용되지 않음

➢ 형사상 적용을 위한 일반적 요구사항:
- 표현에 혐오 선동, 심한 모욕, 극심한 조롱이 있는 경우
- 신체적 위해 혹은 신체적 위해의 협박, 개인이나 집단 혹은 그들의 재산에 대해 물리적 손해를 가하도록 타인을 선동하는 행위

CRICOS code 00025B 4

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

## 민사상 혐오표현 규제법: 모델 #1

연방 인종차별금지법*1975* (Cth), s18C:
사적인 행위 이외의 것으로서 아래와 같은 행위를 하는 것은 불법이다.

상황에 상관 없이, 타인이나 집단에게 불쾌감, 모욕감, 굴욕감을 주거나 위협을 가할 가능성이 상당한 행위
타인 혹은 집단의 일부 혹은 전부에 행해지는 행위의 기준이 그들의 인종, 국적, 종족 출신지 등에 의한 경우 .

'단순 경미한 것이 아닌 극심하고 심각한 영향을 주는 행위'
'선의'에 의한 예술적, 학술적, 과학적, 그리고 언론과 관련된 행위.

CRICOS code 00025B 5

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

## 민사상 혐오표현 규제법 : 모델 #2

뉴사우스웨일즈주 반차별금지법*1977* (NSW):
개인이 공공행위로서, 인종을 이유로 타인 혹은 집단을 특정 지어 행하는 혐오 선동 행위, 심한 모욕, 극심한 조롱은 불법이다..

➢ 예외규정: 공정보도, 선의에 의한 교육, 예술, 과학 혹은 연구 목적

CRICOS code 00025B 6

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

## 민사상 고소 방식

1. 관련 인권 당국에 소장 제출
2. 해당 기관의 기소 내용 조사
3. 해당 기관은 원고와 피고 사이의 기밀 조정을 실시
4. 구제책으로는 행위의 즉시 중지, 사과, 철회 의사 게재 혹은 직장 내에서의 교육이 있음
5. 기소자가 만족스러운 결과를 얻지 못할 경우, 기소 행위를 종결한 후 각 재판소 혹은 법원에서 민사 절차 진행에 착수할 수 있음 (그러나 1.8%의 경우만이 착수)
6. 법원은 구체책으로서 중지명령, 손해배상금 지급, 시정 통지 발행을 명할 수 있음

CRICOS code 00025B 7

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

## **혐오표현 금지법은 얼마나 잘 작동할까?**

1. 구제 방법?
▪ 지지를 위한 기반
▪ 경우에 따라 유리한 결과를 얻어낼 수 있음
▪ 사람들을 차별행위로부터 보호 하려는 정부의 메시지 전달

2. 표현의 수정?
절제된 표현

3. 교육적, 상징적 효과?
• 법의 적용을 받는 사람들의 지지
• 교육적인 방법으로 사용

CRICOS code 00025B 8

4. '냉각효과'?
- 냉각효과에 대한 증거 없음

5. 표현의 자유 '순교자의 양산'?
- 민사상 접근으로 인한 문제 해결:
  - 2% 미만의 경우만 재판소/법원으로 넘어감
  - 민사상 손해배상으로 인한 제재는 가혹하지 않음

CRICOS code 00025B 9

## Lessons

- 사용 가능한 다양한 정의의 존재
- 민법을 강조하면 혐오표현 규제법에 대한 비판을 줄일 수 있음
- 예외규정으로 표현의 자유 보호 가능
- The laws are designed to change *how* people engage in public debate so they do not harm others (not to close off topics of debate) 혐오표현 규제법은 사람들이 대중 토론에 참여하는 방식을 바꿔 다른 사람에게 해를 끼치지 않도록 고안된 것임(토론 주제를 제한시키고자 하는 것이 아님)
- 혐오표현 규제법으로 얻는 이득은 상당하며, 특히 해당 지역 개개인은 물론 사회 공동체 전반에 긍정적 영향을 가져다 줌.
- 혐오발언법으로 인한 부정적 영향은 미미한 수준

CRICOS code 00025B 10

## Scholarly publications

Gelber, K 2019 'Free Speech Debates in Australia – Contemporary Controversies', in H Knowles and B Metroka (eds) *Theorising Free Speech.* Peter Lang (forthcoming).

Gelber, K & McNamara, L 2016 'Anti-vilification laws and public racism in Australia: mapping the gaps between the harms occasioned and the remedies provided', *University of New South Wales Law Journal* 39(2):488-511.

Gelber, K & McNamara, L 2016 'Evidencing the harms of hate speech', *Social Identities*, 22 (1-3): 324-341 (doi: 10.1080/13504630.2015.1128810).

Gelber, K & McNamara, L 2015 'The effects of civil hate speech laws: lessons from Australia', *Law and Society Review* 49(3): 631-664. (DOI: 10.1111/lasr.12152).

CRICOS code 00025B 11

The University of Queensland Australia | CREATE CHANGE

# Thank you

Professor Katharine Gelber
Head, School of Political Science and International Studies
University of Queensland
k.gelber@uq.edu.au

facebook.com/uniofqld
@KGelber

THE UNIVERSITY OF QUEENSLAND AUSTRALIA | CREATE CHANGE

# International approaches to hate speech laws: lessons from Australia

Professor Katharine Gelber
Head of School
School of Political Science and International Studies
University of Queensland
k.gelber@uq.edu.au

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

## Chronology of the introduction of civil hate speech laws

| Jurisdiction | Law enacted | Ground/s added |
|---|---|---|
| **New South Wales** | 1989 | Race – 1989<br>Homosexuality – 1993<br>HIV/AIDS – 1994<br>Transgender – 1996 |
| **Australian Capital Territory** | 1991 | Race – 1991<br>HIV/AIDS – 2004<br>Transsexuality – 2004<br>Sexuality - 2004 |
| **Commonwealth** | 1995 | Race - 1995 |
| **South Australia** | 1996 | Race - 1996 |
| **Tasmania** | 1998 | Race – 1998<br>Religion – 1998<br>Sexuality – 1998<br>Disability - 1998 |
| **Queensland** | 2001 | Race – 2001<br>Religion – 2001<br>Sexuality – 2002<br>Gender identify - 2002 |
| **Victoria** | 2001 | Race – 2001<br>Religion - 2001 |

CRICOS code 00025B 2

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

## Reasons for introducing hate speech laws

1. Existing laws were inadequate to deter public expressions of racism
2. A commitment to multiculturalism
3. A commitment to equality and non-discrimination
4. Obligations under international human rights law
5. Balance between free speech, and the right to a life free from discriminatory harms

- Exemptions:
  - ✓ fair reporting, and
  - ✓ artistic, academic or scientific research, or public debate in 'good faith'
- Emphasis in practice is on civil laws.

CRICOS code 00025B 3

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

## Criminal hate speech laws

- Do not exist in every jurisdiction in Australia
- Almost never used

- Usually require:
  - Speech that 'incites hatred, serious contempt or severe ridicule' **and**
  - Involves physical harm or the threat of harm, or inciting others to threaten physical harm towards a person or group of persons, or property

CRICOS code 00025B 4

## Civil hate speech laws : Model #1

Federal *Race Discrimination Act 1975* (Cth), s18C:

It is unlawful for a person to do an act, otherwise than in private, if:

- the act is reasonably likely, in all the circumstances, to offend, insult, humiliate or intimidate another person or a group of people; and
- the act is done because of the race, color or national or ethnic origin of the other person or of some or all of the people in the group.

'Profound and serious effects, not to be likened to mere slights'

Exemptions: artistic, academic, scientific, and journalistic conduct done in 'good faith'.

CRICOS code 00025B 5

## Civil hate speech laws: Model #2

New South Wales *Anti Discrimination Act 1977* (NSW):

It is unlawful for a person, by a public act, to incite hatred toward, serious contempt for, or severe ridicule of, a person or group of persons on the ground of the race of the person or members of the group.

- Exemptions: fair reporting, done reasonably and in good faith; academic, artistic, scientific or research purposes.

CRICOS code 00025B 6

## Complaint mechanism under the civil law

1. A person lodges a complaint with relevant human rights authority
2. The authority investigates the complaint
3. The authority conciliates a confidential mediation between the complainant and the respondent
4. Remedies can include agreement to stop, apologise, publish a retraction, or conduct an educational campaign in the workplace
5. The complainant, if not satisfied, may terminate complaint and commence civil proceedings in a tribunal or court (but only 1.8% of cases do this)
6. Court ordered remedies can include an order to desist, damages (very small) or the publication of a corrective notice.

CRICOS code 00025B 7

## How well do hate speech laws work?

1. A remedy?
   - A basis for advocacy
   - A favourable outcome for some complaints
   - Government sends a message of protecting people from discrimination

2. Changed speech?
   - In moderated outlets

3. Educative and symbolic?
   - Targetted groups support them
   - They are used in educative ways

CRICOS code 00025B 8

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

4. A 'chilling effect'?

- No evidence of chilling effect

5. Creation of 'free speech martyrs'?

- Reliance on civil laws reduces this problem:
  - Less than 2% of cases go to court/tribunal
  - Civil remedies do **not** have harsh sanctions

CRICOS code 00025B 9

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

## Lessons

- A variety of definitions is available
- An emphasis on civil law reduces many of the criticisms of hate speech law
- Exemptions can protect free speech
- The laws are designed to change *how* people engage in public debate so they do not harm others (not to close off topics of debate)
- Hate speech laws provide benefits to target communities, and through that to the whole community
- Hate speech laws have negligible negative effects

CRICOS code 00025B 10

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

## Scholarly publications

Gelber, K 2019 'Free Speech Debates in Australia – Contemporary Controversies', in H Knowles and B Metroka (eds) *Theorising Free Speech.* Peter Lang (forthcoming).

Gelber, K & McNamara, L 2016 'Anti-vilification laws and public racism in Australia: mapping the gaps between the harms occasioned and the remedies provided', *University of New South Wales Law Journal* 39(2):488-511.

Gelber, K & McNamara, L 2016 'Evidencing the harms of hate speech', *Social Identities*, 22 (1-3): 324-341 (doi: 10.1080/13504630.2015.1128810).

Gelber, K & McNamara, L 2015 'The effects of civil hate speech laws: lessons from Australia', *Law and Society Review* 49(3): 631-664. (DOI: 10.1111/lasr.12152).

CRICOS code 00025B 11

THE UNIVERSITY OF QUEENSLAND AUSTRALIA | CREATE CHANGE

# Thank you

Professor Katharine Gelber
Head, School of Political Science and International Studies
University of Queensland
k.gelber@uq.edu.au

facebook.com/uniofqld
@KGelber

Session 1

# 국내 혐오표현의 유형과 대응방안
# Types and countermeasures of hate speech in South Korea

발제자 | Speaker
류민희 Min-hee Ryu
공익인권변호사모임 희망법 변호사
Lawyer of Korean Lawyers for Public Interest and Human Rights

제5회
국제인권심포지엄
혐오표현의 확산, 어떻게 대응할 것인가?
- 해외 사례로 보는 제도적 대안

# 국내 혐오표현의 유형과 대응방안
# : 2016년 국가인권위원회 혐오표현 실태와 규제방안 실태조사를 중심으로

류민희
공익인권변호사모임 희망을만드는법

## 사회적 문제로서의 혐오표현의 대두

- 한국에서의 인터넷
- 2010년 경 “반다문화” 인터넷 모임 등
- 2012년 경 극우 포퓰리즘 사이트 일베에 대한 사회적 관심
- 혐오/증오범죄로서의 2016년 강남역 살인사건; 가해자의 여성혐오적 표현



## 2016년 국가인권위원회 연구

- 혐오표현이란 “어떤 개인·집단에 대하여 그들이 사회적 소수자로서의 속성을 가졌다는 이유로 그들을 차별·혐오하거나 차별·적의·폭력을 선동하는 표현”으로 정의
- 유형을 △차별적 괴롭힘, △차별표시, △공개적인 멸시·모욕·위협, △증오 선동 등으로 구분
- 온라인 조사 및 대면조사 방법으로 1,014건의 설문을 수집·분석. 여성, 성 소수자, 장애인, 이주민 등에 대한 면접조사를 병행
- 온라인 혐오표현의 실태에 대해서는 별도의 분석

## 결과

- 피해를 입은 소수자집단은 낙인과 편견으로 인해 일과 학업 등 일상생활에서 배제되어 두려움과 슬픔을 느끼고 지속적인 긴장 상태나 무력감에 빠지거나 자존감 손상으로 인한 자살충동, 우울증, 공황발작, 외상 후 스트레스 장애 등에 시달리고 있었음
- 혐오표현을 접한 이후 ‘스트레스나 우울증 등 정신적 어려움을 경험하였다’라는 질문에 장애인(58. 8%), 이주민(56.0%) 성소수자(49.3)% 절반 정도의 응답자가 경험이 있다고 응답

## 대응: 현재의 법적 틀거리

- 형법
  - 명예훼손죄(제307조)와 모욕죄(제311조)
- 민법
  - 일반적인 불법행위 손해배상 청구

## 대응: 현재의 법적 틀거리

- 행정적 규제
  - 성희롱 (국가인권위원회법, 남녀고용평등법 등)
  - 괴롭힘 (장애인차별금지법)
- 온라인 혐오표현, 언론/미디어 상 혐오표현
  - 인터넷 명예훼손 (정보통신망법)
  - 방송통신심의위원회 (방송통신심의위원회 심의규정)
  - 언론 (언론중재 및 피해구제 등에 관한 법률)

## 한계

- 많은 혐오표현은 개인보다는 대상 집단 전체에 관한 것인데, 형사 민사 행정 규제는 개인이나 좁은 범위의 특정 집단에 대한 행위만 다룰 수 있음

## 권장되는 대응방안

- 입법적 대응
  - 혐오표현을 대상으로 한 세밀하고 정확한 입법
  - 현재 제안된 많은 법안들이 명확성을 결여하였거나 대상집단의 포함이 충분하지 않음

## 혐오표현에 대응하고 대응하기

- 모든 행위자는 역할이 있음: 국가인권기구, 입법부, 정부와 지방정부, 사법부, 교육청, 방송통신위원회, 언론, 인터넷중개자, 시민사회
- 하지만 정부의 역할이 강조되어야 함
- “혐오표현은 단지 의견을 표현하는 것에서 그치지 않습니다. 누군가를 때리고, 배제하고, 학교에서 쫓아내거나 직장에서 잘리게 만드는 행위죠. ‘표현의 자유’는 민주주의에서 아주 중요하지만, 분명히 한계는 있습니다. 혐오표현은 그 선을 넘어선 것이고요.” “혐오표현은 정부가 나서서 고심해야만 하는 영역” 빅토르 마드리갈-보를로스 유엔 성소수자 특별보고관

## 권장되는 대응방안

- 근본적 원인을 다루는 긍정적 형성적 정책 방안
  - 표현의 자유의 신장
  - 평등과 반차별 권리의 신장 (포괄적 차별금지법, 국가인권기구/평등기구의 중요성)
  - 공기관의 강력한 입장
    - 공무원 교육
    - 공적 캠페인
    - 시민사회와의 협력/지지



제5회
국제인권심포지엄
혐오표현의 확산, 어떻게 대응할 것인가?
– 해외 사례로 보는 제도적 대안

## 국내 혐오표현의 유형과 대응방안
## Hate speech in Korea and its types and responses: with a focus on 2016 NHRCK study

류민희 공익인권변호사모임 희망을만드는법
Minhee Ryu Korean Lawyers for Public Interest and Human Rights

## Hate Speech as an emerging social problem

- “anti-diversity” internet forum (circa 2010)
- a far-right, right-wing populist website called ILBE gained prominence around 2012
- Gangnam Station Murder in 2016 as a hate crime; the perpetrator manifested misogynic hate speeches

## 2016 NHRCK Study

- The study defined hate speech as "an expression of discrimination, hate, or incitement of discrimination, hostility, or violence against an individual or group because they have attributes as social minorities."
- The study classified the 4 types into discriminative harassment, discrimination signs, public contempt, insults and threats, and incitement of hatred.
- The survey collected 1,014 questionnaires by using online and face-to-face surveys.

## Results

- Victims were excluded from daily life, such as work and schoolwork, because of stigma and prejudice.
- They felt fear and sorrow, suffered from constant tension or helplessness, suicidal impulses from impaired self-esteem, depression, panic attacks, and post-traumatic stress disorder.
- After experiencing hate speech, about half of the respondents responded they felt stressed and depression. [People with disabilities (58.8%), migrants (56.0%), and LGBTI people (49.3%)]

## Current Legal Framework

- Criminal Law
  - Criminalization of defamation and insult (Criminal Act, article 307 and 311)
- Civil Law
  - Tort Claims

## Current Legal Framework

- Administrative Regulation
  - Sexual Harassment (NHRCK act, Act on Equal Employment)
  - Harassment (Disability Discrimination Act)
- Online/Media Hate Speech
  - Online Defamation (Information and Communication Network Act)
  - Korea Communications Standards Commission; Deliberation Regulations on Information and Communication
  - Press Arbitration Law

## Responding and Countering hate speech

- Every actors has a role: NHRI, Legislative Branch, Government and Local governments, Judiciary, Education Office, Communication Standards Commission, Media, Internet intermediaries, Civil Society
- However, we want to stress the role of the government.
- “Hate speech doesn’t just mean expressing an opinion. It’s an act that causes people to be ostracized and beaten, to be kicked out of schools and fired from their jobs. Freedom of expression is very important in a democracy, but there’s obviously a limit to that. Hate speech crosses that line,” Victor-Madrigal Borloz, the UN’s independent expert on protection against violence and discrimination based on sexual orientation and gender identity during the recent interview with a Korean media

## Limitations

- While a lot of hate speech is directed at the entire target group, not individuals, criminal and civil and administrative regulations can deal only with violations of rights to individuals or to specific groups.

## Suggested Responses

- Legislative Responses
  - Nuanced and pin-point measures to target hate speeches
  - Many proposed bills lacked the clarity and inclusion of targeted groups

## Suggested Responses

- Positive Policy Measures to deal with root causes
  - Promotion of freedom of expression
  - Promotion of equality and non-discrimination (comprehensive non-discrimination act, importance of NHRI)
  - Strong position from the public institution
    - Training for public officials
    - Public campaigning
    - Empowering Civil Society